Fault Log Recovery Using an Incomplete-data-trained FDA Classifier for Failure Diagnosis of Engineered Systems

Hyunjae Kim, Jong Moon Ha, Jungho Park, Sunuwe Kim, Keunsu Kim, Beom Chan Jang, Hyunseok Oh and Byeng D. Youn

Department of Mechanical and Aerospace Engineering, Seoul National University, Seoul 151-742, Republic of Korea

secutus07@snu.ac.kr, billyhjm@gmail.com, hihijung@snu.ac.kr, lunashisun@gmail.com keunshu@gmail.com, bob1333@naver.com, hyunseok52@gmail.com, bdyoun@snu.ac.kr

ABSTRACT

In the 2015 PHM Data Challenge Competition, the goal of the competition problem was to diagnose failure of industrial plant systems using incomplete data. The available data consisted of sensor measurements, control reference signals, and fault logs. A detailed description of the plant system of interest was not revealed, and partial fault logs were eliminated from the dataset. This paper presents a fault log recovery method using a machine-learning-based fault classification approach for failure diagnosis. For optimal performance, it was critical to be able to utilize a set of incomplete data and to select relevant features. First, physical interpretation of the given data was performed to select proper features for a fault classifier. Second, Fisher discriminant analysis (FDA) was employed to minimize the effect of outliers in the incomplete data sets. Finally, the type of the missing fault logs and the duration of the corresponding faults were recovered. The proposed approach, based on the use of an incomplete-data-trained FDA classifier, led to the second-highest score in the 2015 PHM Data Challenge Competition.

1. INTRODUCTION

Failure diagnosis of engineered systems plays a critical role in industrial plant systems. A robust and accurate failure diagnosis system helps prevent fatal accidents, saves costs and increases manufacturing efficiency (Hu, Youn, Wang, & Yoon, 2012; Wang, Wang, Youn, & Lee, 2105; Oh, Han, McCluskey, Han, & Youn, 2015). Developing a high-performance failure diagnosis system for a particular system requires mainly two kinds of information: (1) a profound understanding of the target system or (2) condition monitoring / fault log data. An ample level of knowledge about system failures (i.e., mechanisms, root causes) can facilitate effective fault diagnosis for industrial plant systems. On the other hand, a significant amount of monitoring / fault log data – if available – can provide excellent information for data-driven diagnosis (e.g., data mining, machine learning). Unfortunately, having thorough knowledge of the target system is nearly impossible in real plant systems in field applications because such systems are composed of numerous components and operate in a variety of conditions (Kim, Hwang, Park, Oh, & Youn, 2014). Therefore, most fault diagnosis methods focus on securing accurate condition monitoring / fault log data. In reality, however, most available data contains incomplete or missing fault logs due to human factors or monitoring systems that provide poor (e.g., obsolete format) data.

The Prognostics and Health Management (PHM) Society addressed the topic of failure diagnosis of industrial plant systems with incomplete failure log data in the 2015 PHM Data Challenge Competition. The problem in this competition was to identify (1) the types of faults, and (2) the start and end times of the corresponding faults. The problem partially reflects real-world situations because failure logs are often missing in actual real-world industrial applications.

Several approaches for failure diagnosis using incomplete data have been researched (Lee, C., Choi, S. W., Lee, J. M., & Lee, I. B. 2004; Negnevitsky, M. & Pavlovsky, V. 2005; Razavi-Far, R., Zio, E., & Palade, V. 2014; Wu, Y., Jiang, B., Lu, N. Y., & Zhou, Y. 2015). Li et al. (2006) introduced a method for dealing with an incomplete data set using data mining based on rough set theory. In Li's method, a two-stage data mining technique is implemented to extract a diagnostics rule. Li applied the method to a pump system fault diagnosis problem. Marwala and Chakraverty (2006) investigated fault classification in structures with incomplete measured data. They proposed a method based on an autoassociative neural network and a genetic algorithm. First, the neural network is trained with the incomplete data and the genetic algorithm is then used to determine missing input values. Yongli et al. (2006) proposed an approach based on Bayesian networks to deal with uncertain or incomplete data for power system

diagnosis. He et al. (2009) developed robust fault detection for networked systems with communication delay and missing data. He designed a robust fault detection filter for incomplete measurements using H infinity filtering and a Markovian jumping system.

This paper presents the failure diagnosis method used by the SNU-SHRM team and presents the team's results. The key idea for failure diagnosis is to recover the missing fault log data from the industrial plant system of interest. The rest of this paper is organized as follows. Section 2 defines the Data Challenge problem by describing the data set and its structure. Section 3 shows the analysis of the given dataset and extracts the key ideas of the proposed method. The incomplete-data-trained FDA method, along with features that the team suggests for enhancing accuracy, is presented in Section 4. Section 5 presents the results of the fault log recovery. The paper concludes with a summary of the proposed research and suggestions for future work.

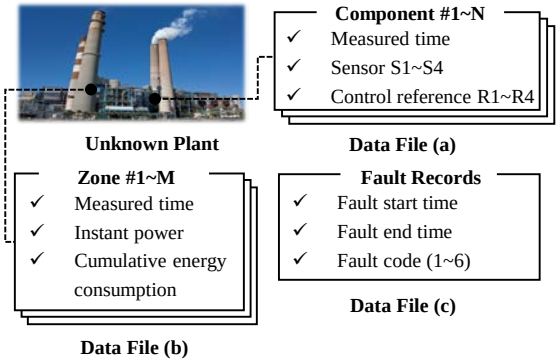


Figure 1. Descriptions of the released data sets

2. PROBLEM AND DATASETS

The problem of the 2015 PHM Data Challenge Competition is described in Section 2.1. The details of the released datasets and the scoring procedures are presented in Sections 2.2 and 2.3, respectively.

2.1. Problem Definition

The goal of the 2015 competition problem is to develop a method that can use incomplete data to (1) determine the type of faults present in the system of interest and (2) predict the start and end times of the faults for unknown industrial plants. The committee provided data sets from 48 plants that included sensor signals and fault logs. Data from 33 plants was complete; however, second half data from the fault logs of 15 plants was partially eliminated in a random manner.

2.2. Description of the Data Sets

As shown in Figure 1, three files for each of 48 plants were released to the participants (Rosca, J., Song Z., Willard, N., & Eklund, N. 2015). Each plant has a different number of components and zones. "File (a)" contains the time series of four sensor signals and four reference signals for N components in that particular plant. The components in a plant are controlled by a feedback loop system. "File (b)" includes the cumulative energy consumption and instantaneous power measured in M zones. "File (c)" contains fault starting times, fault ending times, and fault codes. Each File (c) contains one to six independent fault codes. Among them, fault code 6 is considered trivial as described by the 2015 competition organizer. It is worth noting that the occurrence of any fault is considered to be independent of any other fault. The sensor signals and control references were sampled every 15 minutes. The total time span of data collection of the sample data is approximately three to four years.

2.3. Scoring Process

The score metric is defined as:

$$\text{Score} = 10\times N_{TP} - 0.01\times N_{MS} - 0.1\times N_{FP} - 0.1\times N_{FN} \quad (1)$$

where N_{TP}, N_{MS}, N_{FP}, and N_{FN} are the number of true positives (TP), misclassifications (MS), false positives (FP), and false negatives (FN), respectively.

The score varies with the number of correct or false predictions. The scoring system awards ten points for true positives. If the fault prediction is placed within the one-hour tolerance of the actual fault time and has the correct fault code, the prediction is accepted as a true positive. The scoring system penalizes misclassifications, false positives, and false negatives. A misclassification corresponds to faults identified with correct start and end times, but with the wrong fault code. False positives indicate faults identified by a submission that did not actually occur in the data. False negatives mean faults in the actual data that are not identified in the submission.

3. DATA ANALYSIS

This section investigates the characteristics of the given data in such a way that the findings in the section can be used for fault log recovery, as proposed in Section 4. The correlation between sensor measurements and reference control signals is identified in Section 3.1. To define the dataset for training a classifier, seasonality analysis is presented in Section 3.2. In Section 3.3, statistical analysis of sensor signals and fault data is shown to identify the distribution of fault durations. In Section 3.4, rule-based fault diagnostics is presented to verify the applicability of machine-learning-based fault log recovery.

3.1. Sensor Data Analysis based on Inference of the System Type from a Physical Interpretation

Specifications and details about the industrial plants were not revealed. Therefore, it was impossible to identify the characteristics of the exact system and the collected data. However, there were some clues from which we could infer

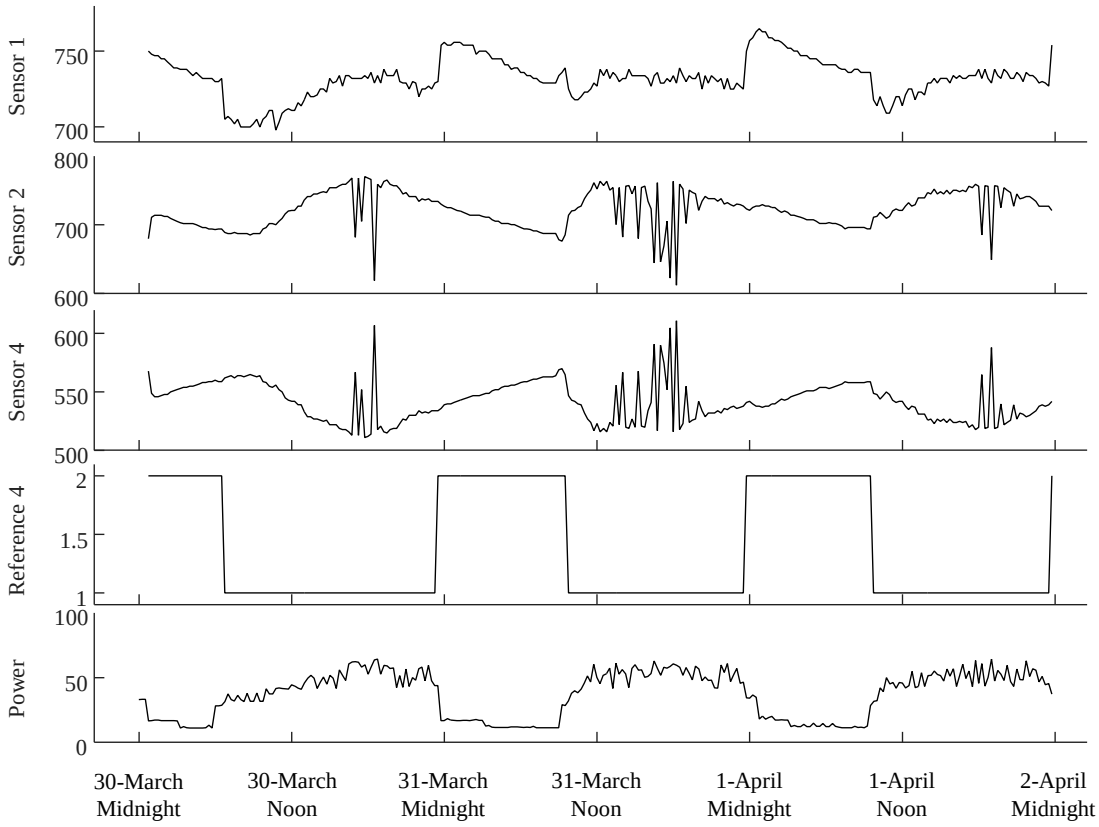


Figure 2. Sensor and reference control data trends for three days

the type of system. For example, the terms used to describe the problem, including 'sensor signal,' 'control reference,' and 'energy consumption' provided us keywords for a literature review. Thus, we attempted to find a plant system with similar terminology, signals, and operating modes, as described below.

A correlation between sensor measurements and reference control signals was observed in studies of air handling units from heating, ventilation, air conditioning (HVAC) systems, as discussed by Schein (2006). For example, in Figure 2, we recognized that signals from Sensors 2 and 4, which can be related to those from heating and cooling sensors of air handling units (AHU), show behaviors opposite to each other. When the amplitude from Sensor 2 rises, that of Sensor 4 falls. The correlation coefficient between them was almost minus one. Thus, we assumed that these sensor values had an inversely-proportional relationship.

Furthermore, the reference signals operated in a way in which they controlled the valve position or pre-determined temperature, as described by Salsbury (2001). The value of the reference signal usually changed periodically both day and night. From this, we suspected that Sensor 1 could be an object value of the system. In the daytime, it was observed that the magnitude from Sensor 1 fluctuated after that from Reference 4 changed. The instantaneous power consumption value was also related to signals from Sensor 1 and Sensor 2. The value of Sensor 1 has a trend that approximately correlates the Sensor 2 value, especially working time. However, this trend does not hold at the moment of transition from the working time to night time. From this evidence, it is reasonable to assume that Sensor 1 shows the target temperature. Sensor 2 is a representative value of the heating operation function. Sensor 4 indicates the cooling operation function in the temperature regulation system. These findings are used to extract proper features, as outlined in Section 4.2.

3.2. Seasonality Analysis for Sensor and Zone Data along with Fault Data

In Section 3.1, we inferred the type of the system of interest from the perspective of the physical mechanisms. This section investigates the characteristics of a time series in sensors and fault logs that recurs every calendar year. Figure 3 shows a representative example including data from Sensors 1, 2, and 3, and data on instantaneous power for two years. Both sensor magnitudes and instantaneous power have a local minimum in winter, and a local maximum in summer. This predictable pattern existed over a one-year period. For

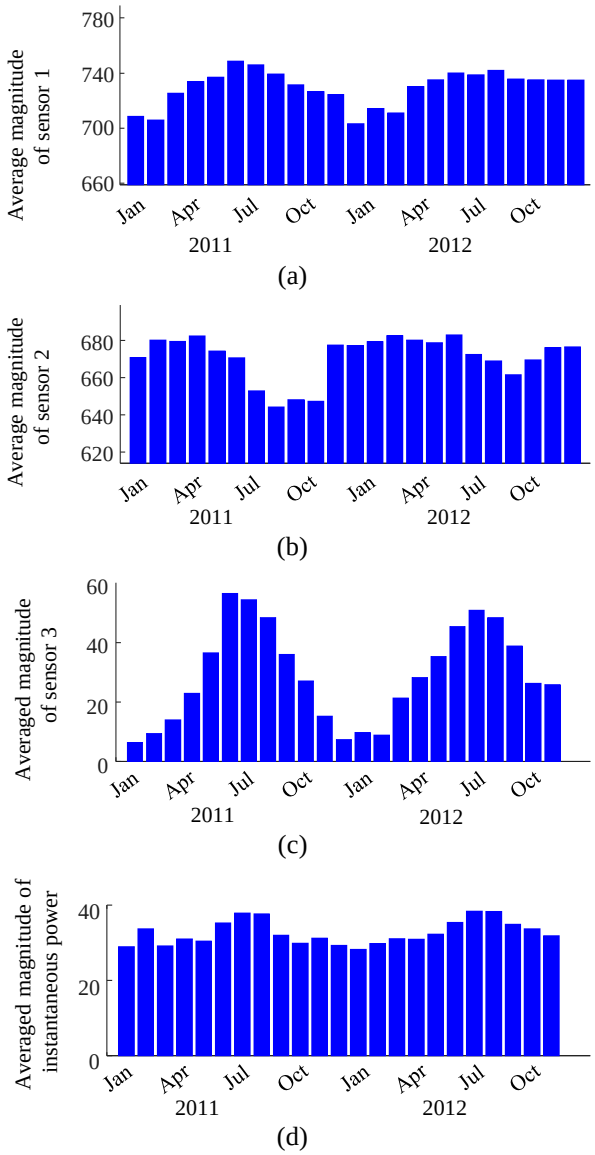


Figure 3. Average magnitudes and zone instantaneous power for a two-year period

Table 1. Fault frequency for three years, by quarter

		Fault Code 1	Fault Code 2	Fault Code 3	Fault Code 4	Fault Code 5
2010	Q1	96	20	0	0	0
	Q2	121	56	**0**	26	0
	Q3	87	87	1	28	0
	Q4	48	0	0	0	0
2011	Q1	0	0	0	0	0
	Q2	34	179	**29**	44	58
	Q3	78	55	30	18	25
	Q4	13	5	0	0	5
2012	Q1	1	3	0	0	0
	Q2	140	78	**0**	0	0
	Q3	346	64	0	0	0
	Q4	69	0	0	0	0

this reason, it was assumed that the industrial plant data was from a temperature controlling system.

Table 1 summarizes the number of faults for each quarter over three years. Despite the seasonal characteristics in the sensor and zone data, the occurrence of faults does not show a seasonality pattern. Even though some faults frequently occurred in a particular quarter of a year (e.g., F3 in Q2 of 2011), these same faults may not be found in the same quarter of the next year (e.g., F3 in Q2 of 2012). Meanwhile, the opposite situation can also be observed. Even though no fault occurred in one quarter of a particular year (e.g., F3 in Q2 of 2010), a fault may be found in the same quarter of the next year (e.g., F3 in Q2 of 2011).

Based on observations of the seasonality analysis, it was identified that the occurrence of faults does not show seasonality, whereas the sensor data does. This indicates that use of the sensor data from the first year and the corresponding fault logs for training purposes may not be the best solution to accurately detect faults in the second year. The selection of relevant sensor and fault data for training is critical. In Section 4.3, we present our proposed strategy for designing and training a relevant classifier using an incomplete dataset.

3.3. Fault Duration Analysis

Analysis of the duration of faults can provide information about the general characteristic behaviors of a faulty condition in a plant. Figure 4 shows the fault duration time for all logged faults contained in the data from the 33 training plants—those with complete data. Similar to the sensor signals sampled every 15 minutes, fault times were also resampled at 15-minute intervals. Figure 4 shows that 15-minute and 60-minute-long faults occur most frequently, accounting for 25% and 12% of all faults, respectively. 77% of all faults were 180 minutes or less in duration, and only 1.45% of all faults lasted longer than a day.

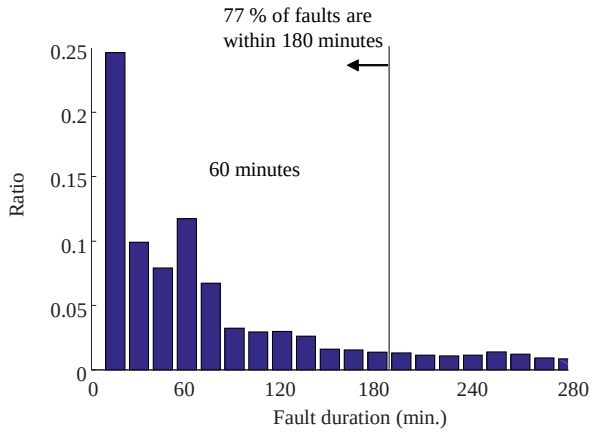


Figure 4. Fault duration times observed in the sample data

3.4. Statistical Analysis for Verification of Applicability of Machine-Learning-based Fault Log Recovery

In the data-driven approach, the basic assumption for fault classification is that a detectable change in health conditions can be observed from a system of interest. Based on the assumption, the empirical PDF of sensor signals can help distinguish normal conditions from faulty conditions. Using the data collected from sensor signals, empirical PDFs are compared in Figure 5. S1, S2, and P in the left side of the figure indicate the data from Sensors 1 and 2 and the average of the instantaneous power, respectively. The indicators in the first row of the figure represent the number of the plant and of the component, respectively. For example, "P10/C1" means that component one in plant 10 was used for the statistical analysis. The distributions with blue and red colors correspond to normal and faulty conditions, respectively. Visual inspection shows that empirical PDF for sensor signal data from normal and faulty conditions are partially separated in some examples. In the highlighted box shown in Figure 5, for example, normal and faulty conditions in plant 10 were partially separable in terms of data from Sensor 2 and the average of the instantaneous power. In this case, most data under the faulty condition had a Sensor 2 value smaller than 470, and an average of instantaneous power larger than 60.

When the time series of the signal from Sensor 2 and the average of the instantaneous power of plant 10 are analyzed, relevant features can be more clearly found, as shown in Figure 6. Data for "Fault Code 5" is marked with red circles in Figure 6. In this example, the fault was detected around one hour before the following simple rules were satisfied: (1) Sensor 2 data was lower than 470, and (2) the average of the instantaneous power in the zone was greater than 60. Based on the abovementioned rules, a rule-based fault detection was attempted for Fault Code 5 of plant 10, as shown in Figure 7. The faults predicted by the rule are represented by blue cross marks in the figure. As a result, 86.28% of Fault Code 5 events in plant 10 were successfully predicted.

Although the abovementioned rule-based fault diagnostics approach successfully identified the existence of Fault Code 5 in plant 10, it was not generally applicable for other plants. Most plants have an irregular number of components and zones. As a result, there may exist hundreds of rules that define faults of these systems based on the combination of multiple signals from several components and zones. In this case, it is impossible to identify general rules for diagnostics of most faults. Because of this challenge, the following section of this paper presents a machine-learning-based fault log recovery that can substitute for rule-based fault diagnostics.

4. FAULT LOG RECOVERY FOR FAILURE DIAGNOSIS

In this section, a fault log recovery technique is proposed for failure diagnosis. Relevant features are extracted based on physical interpretation of the data. Then, an FDA-based classifier is proposed to incorporate incomplete data. Finally, the procedures of fault log recovery for failure diagnosis are presented.

4.1. Processing Fault Logs for Machine Learning

The fault logs in the original file consist of start and end times for each fault. We discretized fault log data every 15 minutes to run a machine-learning algorithm. After processing the fault logs in this manner, the logs have discrete values every 15 minutes. For example, for a fault log that starts at 1:00 pm and ends at 1:45 pm, the log is converted into four discrete fault logs corresponding to fault analysis at 1:00 PM, 1:15 PM, 1:30 PM, and 1:45 PM. This process makes fault logs match signal data.

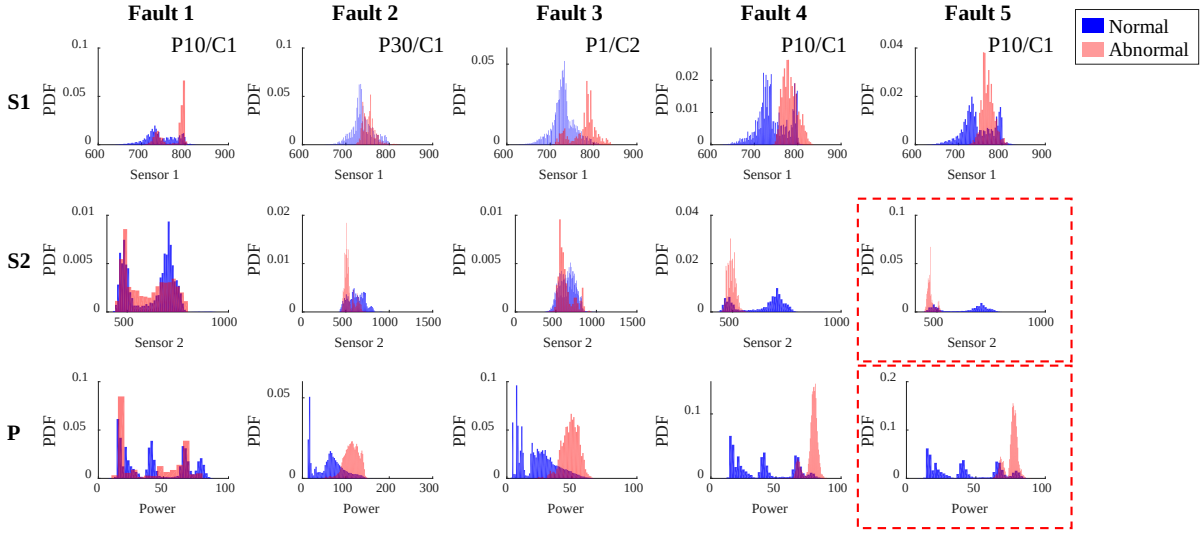
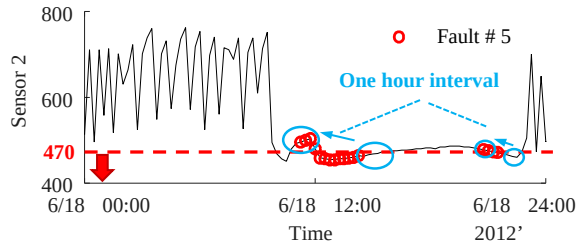
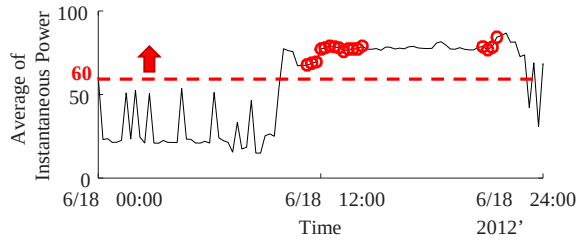


Figure 5. Comparison of empirical PDF for normal and faulty data

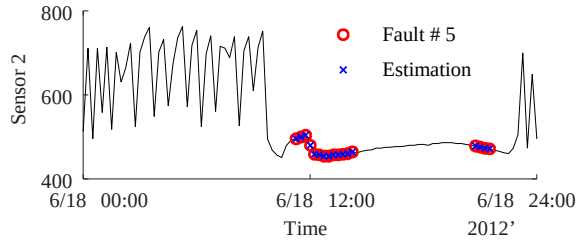


(a)

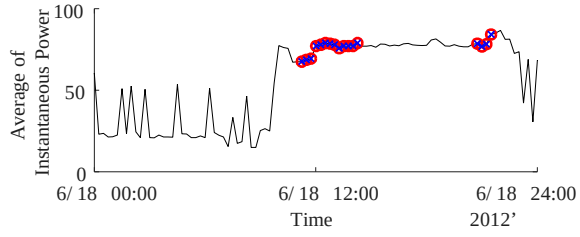


(b)

Figure 6. Time series: (a) data from Sensor 2 and (b) average of the instantaneous power



(a)



(b)

Figure 7. Rule-based fault detection for Fault Code 5 of planet 10: (a) data from Sensor 2 and (b) average of instantaneous power in the zone

4.2. Feature Extraction based on Physical Interpretation of Datasets

In this section, we introduce features in conjunction with the physical interpretation of sensor signals and fault data. As discussed in Section 3.1, we observed that Reference 4 values were discretized with two values, one and two. The two values repeatedly occur during the day and night. In the given data, more faults occurred during the day, when Reference 4 values are one. Therefore, the relationship between sensor values and faults could be enhanced by multiplying each sensor value by the Reference 4 value, as:

$$F_{Si,R4} = S_i \times R_4 \ (i = 1, 2, 3, \text{and } 4) \quad (2)$$

Most instances of Fault Codes 2 to 5 happen when Sensor 1 and instantaneous power are high, as mentioned in Section 3.2. This implies that a faulty condition can be separated from a normal condition if data from Sensor 1 and instantaneous power are integrated into a single feature. The ratio of Sensor 1 to the instantaneous power is defined as:

$$F_{S1,P} = S_1 / P_{inst} \quad (3)$$

From (2) and (3), the features, $F_{S1,R4}$, $F_{S2,R4}$, $F_{S3,R4}$, $F_{S4,R4}$, and $F_{S1,P}$ at time t, are shown in a vector form:

$$F_t = [F_{S1,R4}, F_{S2,R4}, F_{S3,R4}, F_{S4,R4}, F_{S1,P}] \quad (4)$$

Equation (4) shows the features at time t. Features at t-15, t-30, … can be also be presented. In this study, to incorporate the features from the past three hours (180 minutes), the features are stacked like this:

$$F_{stack,t} = [F_t, F_{t-15}, ..., F_{t-180}] \quad (5)$$

F_t consists of five components. Thirteen feature vectors at t, t-15, …, t-180, thus becomes a 1 by 65 matrix, as illustrated in Figure 8.

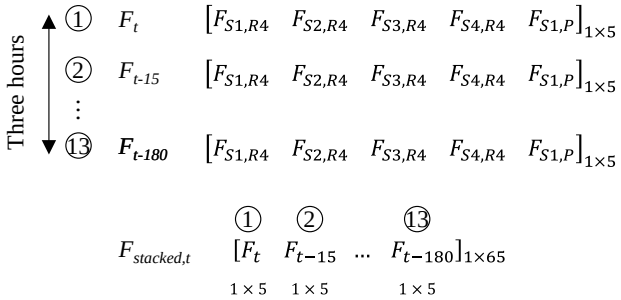


Figure 8. Concept of stacked features

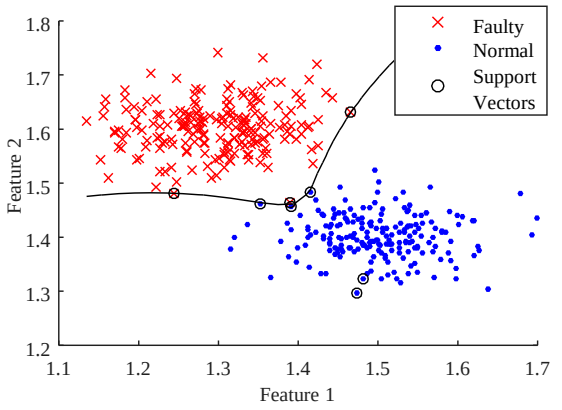
4.3. Incomplete-data-trained Fisher Discriminant Analysis

We propose an incomplete-data-trained FDA for fault data recovery. This method is distinguished from how FDA is usually trained in that the training set contains only partial data (i.e., the second half of the data was removed, as was the case with the data from 15 of the plants). From the viewpoint of a machine learning technique, missing fault log data is interpreted as mislabeled normal data, which is originally faulty. Generally, this mislabeled data is treated like an "outlier" of the data distribution. Because of this, when an

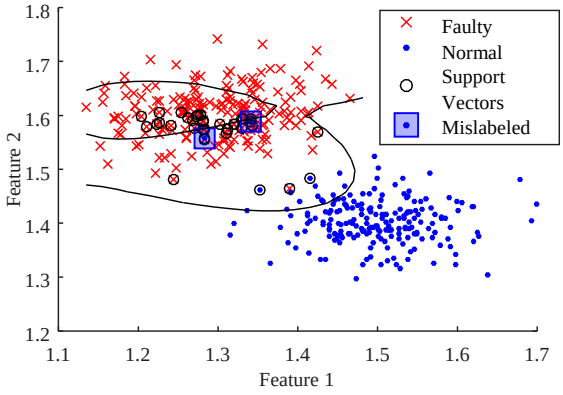
incomplete dataset with mislabeled data is used for training, the accuracy of the fault classification will be lower than when using a complete data set.

The use of an irrelevant classifier has more impact on data points with incorrect labels than for those with correct labels. For example, in Figure 9, a widely accepted classifier like support vector machine (SVM) does not provide high classification accuracy when used with the incomplete dataset. The SVM is designed to find a hyperplane that has a good separation ability by making the hyperplane with the largest margin to the nearest training data for each label. With an incomplete dataset, there was the possibility that the support vector was misplaced due to several mislabeled data points. Thus, the hyperplane did not separate the normal conditions from faulty conditions.

Similarly, another popular memory-based learning technique, k-nearest neighbors (KNN) is also not appropriate for training with this incomplete dataset. KNN has a shortcoming in this setting because it is very sensitive to the data's local structure. If a single mislabeled data point exists in the middle of another label, most of the classification



(a)



(b)

Figure 9. Fault classification by radial basis SVM:
(a) trained using a complete dataset
(b) trained using an incomplete dataset

results near the mislabeled data are labeled to the wrong one. Therefore, a classifier insensitive to the existence of mislabeled data should be identified for use in settings like this one, where incomplete data training is needed.

Unlike these previously described classifiers, the Fisher discriminant analysis (FDA) classifier has robust characteristics for working with incomplete data. FDA can classify normal and faulty data while ignoring a small number of outliers, i.e., mislabeled data (Jeon, Jung, Youn, Kim, & Bae, 2015). This characteristic relies on two facts. First, FDA requires a training set in which only a small portion of the dataset describes faulty conditions. In other words, the training data must consist of a significant amount of normal data and a relatively small amount of faulty data. Second, FDA chooses its separation plane based on each label group's mean and variance. Thus, even if some faulty data are mislabeled as normal, the mean and variance values do not change much. We believe that this characteristic makes the FDA classifier the most suitable classifier for the given incomplete dataset. Figure 10 shows the robustness of FDA to the incomplete dataset. As shown in Figure 10, the value of the separation plane and the classification accuracy trained from the complete dataset and that derived from the incomplete dataset with a few mislabeled data was almost identical. The separation plane does not change in any significant way between the first and second case, while keeping the level of accuracy.

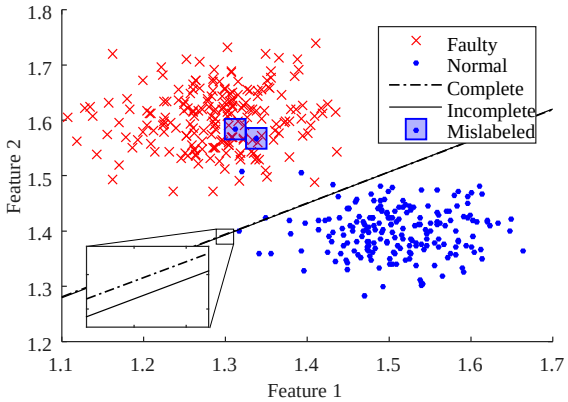


Figure 10. Fault classification by FDA: the results with complete and incomplete datasets are almost identical.

4.4. Procedures of Fault Log Recovery for Failure Diagnosis

The procedures for fault log recovery consist of three steps, as shown in Figure 11: (1) feature values are calculated using the training data set. The selection of a proper training data set enhances the separability of the FDA classifier; (2) a trained FDA classifies normal and faulty conditions for the test data set. In this step, incomplete data missing from the fault logs are recovered through the characteristics of FDA and datasets; (3) the adjacent logs are merged and then

converted back into the original log form and divided into one hour units, as explained in Section 3.3. The fault log recovery technique can also be used in real-time condition monitoring to diagnose failure of plant systems.

5. RESULTS AND DISCUSSION

To validate the proposed method, we randomly eliminated half of the faults in the 33 plants with full fault logs, and evaluated the performance of the method using the score metric in Equation (1). On average, it scored 1663 points per plant with 213 TPs and 4483 FPs per plant. 427 faults were eliminated from the plant fault logs and the method correctly recovered about half of the faults for the individual plants. Although FP values are about 20 times larger than TP values, the FP has a less significant impact on the score than does the TP value.

We incorporated the data from the second half of the data (the data with missing fault logs) to recover the incomplete data as well as other data from various plants. It should be noted that the proposed method can be used for fault log recovery of any industrial plant system, and eventually, for failure diagnosis using real-time condition monitoring data.

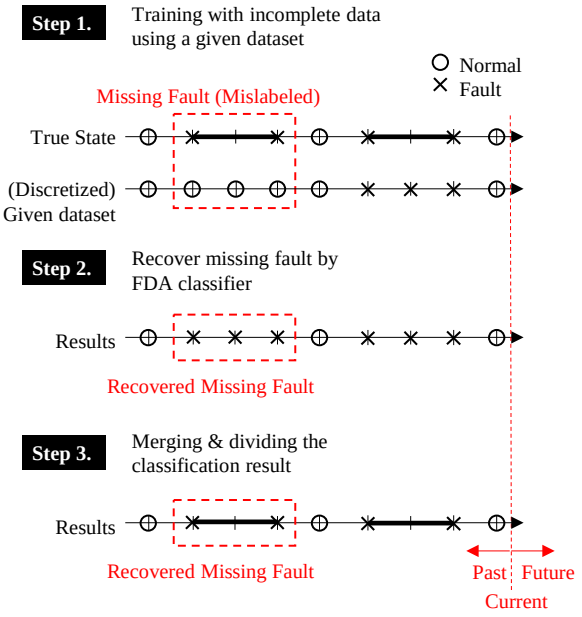


Figure 11. Overview of Fault Log Recovery

Table 2. Scoring details of different classification methods

	TP / plant	FP / plant	Score / plant
Submitted	220	3941	1792
Different training region (first half)	177	5164	1228
Different feature (non-stacked)	224	6253	1591

6. CONCLUSIONS

This study addressed failure diagnosis of industrial plant systems in real applications. The key idea was to recover missing data from incomplete fault logs. The recovery of missing fault data was accomplished through comprehensive analysis of sensor measurements, control reference signals, and fault log data. Data analysis provided correlation between sensor signals and fault logs. A strategy was proposed to recover the missing fault log information and, thus, enable the use of incomplete training data. Compared to other classifiers such as SVM and KNN, the incomplete-data-trained FDA classifier was superior at classifying normal and faulty conditions. The results from the selected features and the FDA-based fault classification method ranked second-highest in the 2015 PHM Data Challenge Competition.

There is a room for further improvement of the proposed method. It would be possible to improve the fault log recovery performance by optimally combining the first and second half of the datasets for use in training the FDA. In addition, it is expected that greater accuracy would be accomplished if more system details become available so that physical interpretation-based features could be defined.

REFERENCES

He, X., Wang, Z., & Zhou, D. H. (2009). Robust fault detection for networked systems with communication delay and data missing. Automatica, 45(11), 2634-2639.

Hu, C., Youn, B. D., Wang, P., & Yoon, J. T. (2012). Ensemble of data-driven prognostic algorithms for robust prediction of remaining useful life. Reliability Engineering and System Safety, 103, 120-135.

Jeon, B. C., Jung, J. H., Youn, B. D., Kim, Y.-W., & Bae, Y.-C. (2015). Datum unit optimization for robustness of a journal bearing diagnosis system. International Journal of Precision Engineering and Manufacturing, 16(11), 2411-2425.

Rosca, J., Song Z., Willard, N., & Eklund, N. (2015). PHM15 challenge competition and data set: fault prognostics, *NASA Ames Prognostics Data Repository*, NASA Ames Research Center, Moffett Field, CA. http://ti.arc.nasa.gov/project/prognostic-data-repository

Kim, H., Hwang, T., Park, J., Oh, H., & Youn, B. D. (2014). Risk prediction of engineering assets: An ensemble of part lifespan calculation and usage classification methods. International Journal of Prognostics and Health Management, 5(2), 1-7.

Lee, C., Choi, S. W., Lee, J. M., & Lee, I. B. (2004). Sensor fault identification in MSPM using reconstructed monitoring statistics. Industrial & Engineering Chemistry Research, 43(15), 4293-4304.

Li, J. R., Khoo, L. P., & Tor, S. B. (2006). RMINE: a rough set based data mining prototype for the reasoning of incomplete data in condition-based fault diagnosis. Journal of Intelligent Manufacturing, 17(1), 163-176.

Marwala, T., & Chakraverty, S. (2006). Fault classification in structures with incomplete measured data using autoassociative neural networks and genetic algorithm. Current Science-Bangalore, 90(4), 542.

Negnevitsky, M. & Pavlovsky, V. (2005). Neural networks approach to online identification of multiple failures of protection systems. IEEE Transactions on Power Delivery, 20(2), 588-594.

Oh, H., Han, B., McCluskey, P., Han, C., & Youn, B. D. (2015). Physics-of-failure, condition monitoring, and prognostics of insulated gate bipolar transistor modules: A review. IEEE Transactions on Power Electronics, 30(5), 2413-2426.

Razavi-Far, R., Zio, E., & Palade, V. (2014). Efficient residuals pre-processing for diagnosing multi-class faults in a doubly fed induction generator, under missing data scenarios. Expert Systems with Applications, 41(14), 6386-6399.

Salsbury, T. I., & Diamond, R. C. (2001). Fault detection in HVAC systems using model-based feedforward control. Energy and Buildings, 33(4), 403-415.

Schein, J., Bushby, S. T., Castro, N. S., & House, J. M. (2006). A rule-based fault detection method for air handling units. Energy and Buildings, 38(12), 1485-1492.

Wang, P., Wang, Z., Youn, B. D., & Lee, S. (2105). Reliability-based robust design of smart sensing systems for failure diagnostics using piezoelectric materials. Computers & Structures, 156, 110-121.

Wu, Y., Jiang, B., Lu, N. Y., & Zhou, Y. (2015). Bayesian network based fault prognosis via Bond graph modeling of high-speed railway traction device. Mathematical Problems in Engineering, 2015.

Yongli, Z., Limin, H., & Jinling, L. (2006). Bayesian networks-based approach for power systems fault diagnosis. IEEE Transactions on Power Delivery, 21(2), 634-639.

BIOGRAPHIES

Hyunjae Kim received his B.S. degree from Seoul National University, Seoul, Republic of Korea, in 2012. He is a Ph.D. student in Seoul National University. His research topic is battery thermal and power management. He received two awards including the IEEE PHM Data Challenge Competition Winner (2014) and the PHM Society Data Challenge Competition Winner (2014).

Jong Moon Ha received his B.S. degree from Hongik University, Seoul, Republic of Korea, in 2011. He a Ph.D. student at the Department of Mechanical and Aerospace Engineering in Seoul National University. His research topic is prognostics and health management for gearboxes. He received an award including the IEEE PHM Data Challenge Competition Winner (2014).

Jungho Park received his B.S. degree from Seoul National University, Seoul, Republic of Korea, in 2012. He is a Ph.D. student at the Department of Mechanical and Aerospace Engineering in Seoul National University. His research topic is fault diagnostics of planetary gears. He received an award including the PHM Society Data Challenge Competition Winner (2014).

Sunuwe Kim received his B.S. degree from Korea University, Seoul, Republic of Korea, in 2014. He is an M.S. student at the Department of Mechanical and Aerospace Engineering in Seoul National University. His research topic is prognostics and health management for lithium-ion batteries.

Keunsu Kim received his B.S. degree from Seoul National University, Seoul, Republic of Korea, in 2013. He is a Ph.D. student at the Department of Mechanical and Aerospace Engineering in Seoul National University. His research area includes prognostics and health management for machine tools. He received an award including the IEEE PHM Data Challenge Competition Winner (2014).

Beom Chan Jang received his B.S. degree from Seoul National University, Seoul, Republic of Korea, in 2014. He is a Ph.D. student at the Department of Mechanical and Aerospace Engineering in Seoul National University. His research topic is prognostics and health management of lithium-ion batteries.

Hyunseok Oh is received the B.S. degree from Korea University, Seoul, Republic of Korea, in 2004, the M.S. degree from KAIST, Daejeon, Republic of Korea, in 2006, and the Ph.D. degree from the University of Maryland, College Park, MD, USA, in 2012. He is a Research Professor in the Laboratory for System Health & Risk Management, Seoul National University. His current research area includes prognostics and health management and model verification and validation. He was with the Hyundai MOBIS Technical Research Institute as a Research Engineer from 2006 to 2007. He was a Research Associate at the Center for Advanced Life Cycle Engineering, University of Maryland from 2012 to 2014. Dr. Oh received the A. James Clark Fellowship (2007). He received two awards including the IEEE PHM Data Challenge Competition Winner (2012) and the PHM Society Data Challenge Competition Winner (2014).

Byeng D. Youn received the B.S. degree from Inha University, Incheon, South Korea, in 1996, the M.S. degree from KAIST, Daejeon, Republic of Korea, in 1998, and the Ph.D. degree from the University of Iowa, Iowa City, IA, USA, in 2001. He is an Associate Professor of mechanical and aerospace engineering at Seoul National University (SNU), Seoul, Republic of Korea. Before joining SNU, he was an Assistant Professor in the Department of Mechanical Engineering, University of Maryland, College Park. His research goal is to develop rational reliability and design methods based on mathematics, physics, and statistics for use in complex engineered systems, mainly focused on energy systems. His current research includes reliability-based

design, prognostics and health management (PHM), energy harvester design, and virtual product testing. Dr. Youn's dedication and efforts in research have garnered substantive peer recognition resulting in four notable awards including the ASME IDETC Best Paper Awards (2001 and 2008), the ISSMO/Springer Prize for a Young Scientist (2005), the IEEE PHM Competition Winner (2014), the PHM society Data Challenge Competition Winner (2014), etc.